<u>Example</u> revisited: $Y_1, \dots, Y_n \sim \text{Geometric}(p)$.

$$\mathcal{L} = \prod_i \mathcal{L}(Y_i | p) = \prod (1-p)^{y_i - 1} p = p^n (1-p)^{\Sigma y_i - n} = \left(\frac{p}{1-p}\right)^n (1-p)^{\Sigma y_i}$$

Write $y = \Sigma y_i$.

$$\ell = n \log p - n \log(1-p) + y \log(1-p)$$

$$\frac{d\ell}{dp} = \frac{n}{p} - \frac{n}{1-p} \cdot (-1) + \frac{y}{1-p} \cdot (-1) = \frac{n}{p} + \frac{n}{1-p} - \frac{y}{1-p} =$$

$$= \frac{np + n(1-p) - yp}{p(1-p)} = \frac{n - yp}{p(1-p)}$$

$= 0$ if $n - y\hat{p} = 0$ if $n = y\hat{p}$ if $\hat{p} = \frac{n}{y} = \frac{n}{\Sigma y_i} = \frac{1}{\bar{Y}}$.

<u>Note</u> On April 29, we will have our lecture together with the Topology course; we will talk about my research field: topological data analysis.

Hypothesis Testing

So far we have been doing <u>estimation</u>
"how large is X?"

The next main task in statistics is <u>hypothesis testing</u>:
"is it true that X?"

The usual way to set this up nowadays is by pitting a "null hypothesis" against an "alternative hypothesis".

When starting from a research hypothesis, it is traditional to set the alternative to match the research hypothesis — and let the null be its converse.

In the framework of statistical testing, a <u>hypothesis</u> is a subset of the parameter space.

Very common shapes of the hypothesis space are

H_0	H_a	
$\vartheta \le \vartheta_0$	$\vartheta > \vartheta_0$	upper tail
$\vartheta \ge \vartheta_0$	$\vartheta < \vartheta_0$	lower tail
$\vartheta = \vartheta_0$	$\vartheta \ne \vartheta_0$	two tailed

With the two competing hypotheses established, the next step is to pick some test statistic — a statistic that will form the basis of our decision to reject the null or not.

The space of possible values for the test statistic is divided into components:

the rejection region (RR) — all values of the test statistic that cause us to reject H_0

since rejection is what interests us, the complement — the "acceptance region" or "non-rejection region" usually doesn't merit notation of its own.

We distinguish two types of possible errors when testing:

Type I / False rejection / False discovery / False positive
Rejecting H_0 when H_0 is true

Type II / False acceptance / False non-discovery / False negative
Failing to reject H_0 when H_A is true

Using these, two probabilities are defined, fundamental to describing hypothesis tests:

$\alpha = \mathbb{P}(\text{false rejection})$ the level of the test

$\beta = \mathbb{P}(\text{false acceptance})$ the miss rate of the test

These are two quantities in a large and useful family of confusion matrix descriptors

The confusion matrix of classifying is a matrix with true classes as columns, predicted classes as rows, and a count in each entry.

Here we get:

		H_A: True	H_A: False
Test	rejects H_0	TP	FP (Type I)
Test	accepts H_0	FN (Type II)	TN

TP	True positive
FP	False positive
TN	True negative
FN	False negative

$TP/(TP+FN)$	Recall / Sensitivity / Power
$FN/(TP+FN)$	β / miss rate
$FP/(FP+TN)$	α / false positive rate / fallout
$TN/(FP+TN)$	specificity / selectivity
$(TP+TN)/N$	accuracy
$TP/(TP+FP)$	precision
$FP/(TP+FP)$	false discovery rate

Even though in the one-tailed tests, H_0 is an entire range of values, it is very common to give H_0 as just a single value: $H_0 : \vartheta = \vartheta_0$ vs $H_A : \vartheta > \vartheta_0$.

Since we are "trying" to reject H_0, the distinction turns out not to matter much.

α as a quantity only really depends on this ϑ_0.
β depends on what the truth is considered to be: strictly speaking, β is a function from H_A to probability values; takes a parameter value from H_A and produces a probability of being wrong in that exact case.

Hence, the <u>power</u> $1-\beta$ is usually analyzed as a function, the <u>power curve</u>.

If we enlarge the RR, then we reject more often, so we reject falsely more often: α increases, β decreases.
If we shrink the RR, then we accept more often. α decreases, β increases.

We might not be able to get exactly the α we want to pick:

<u>Example</u> (Ex. 10.1)

$n=15$ voters were sampled and asked if they favor candidate X.
We are testing $H_0: p=0.5$ against $H_A: p<0.5$.

RR cutoff	1	2	3	4	5	6
α	.0005	.0037	.0175	.059	.151	.303

$$\alpha = P(Y \le \text{cutoff} \mid p=0.5) = \sum_{y=0}^{\text{cutoff}} \binom{15}{y} \cdot 0.5^{15}$$

There is <u>no</u> cutoff we can pick that gives an α of exactly .05.

Randomized Testing

One solution is to introduce randomness to the testing process: we do this by creating a <u>test function</u> $\varphi(y)$ that for each value of the test statistic produces a <u>rejection chance</u>

A non-randomized test is one with $\varphi(y) = \mathbb{1}_{RR}(y)$.

We define $\alpha = \sup\limits_{\vartheta \in H_0} \mathbb{E}[\varphi(x) \mid \vartheta]$

For the example here, $\varphi(1)=\varphi(2)=\varphi(3)=1$; $\varphi(5 \text{ or higher})=0$ and $\varphi(4)$ is chosen to get $\alpha = 0.05$

In other words,

$$0.05 = \alpha = 0.5^{15} \cdot \left(\binom{15}{0} + \binom{15}{1} + \binom{15}{2} + \binom{15}{3} + \varphi(4) \binom{15}{4} \right)$$

$$= 0.5^{15} \cdot \left(1 + 15 + 105 + 455 + \varphi(4) \cdot 1365 \right)$$

So

$$\varphi(4) \cdot 1365 = \frac{0.05}{0.5^{15}} - 576$$

$$\varphi(4) = \left(\frac{0.05}{0.5^{15}} - 576 \right) / 1365 \approx 0.78.$$